ITU - Telecommunications Standardization Sector
STUDY GROUP 16 Question 6
Video Coding Experts Group (VCEG)

15th Meeting: Pattaya, Thailand, 4-6 Dec., 2001

Document VCEG-O37
Filename: VCEG-O37.doc
Generated: 26 November '01

Question: Q.6/SG16 (VCEG)

Source: L. Wang, K. Panusopone, R. Gandhi, Y. Yu and A. Luthra
Motorola Inc.
Broadband Communications Sector
6450 Sequence Drive
San Diego, CA 92121
USA

Tel: 858-404-3893
Fax: 858-404-2501
Email: liwang@gi.com

Title: **Interlace Coding Tools for H.26L Video Coding**

Purpose: Proposal

Abstract

The document describes the possible interlace coding tools for H.26L video coding, and reports the computer simulation results for interlace video sequences as well as the comparisons with the current H.26L TML8.5. It is shown that by using Interlace Coding syntax and tools the performance of TML8.5 can be increased significantly.

1. Introduction

H.26L TML 8.5 [1] allows three possible picture types for an input frame: I, P and B. A frame of any picture type I, P or B is divided into MBs of 16x16 pixels. Each MB of 16x16 can be further divided into blocks in one of seven patterns (modes), as shown in Fig. 1. Block size can be 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, or 4x4. A MB can be coded in either intra or inter. For intra mode, MB can be only in either mode 1 (block size of 16x16) or mode 7 (block size of 4x4). For inter mode, MB can be in any of seven modes. Motion estimation and compensation (ME/MC) is performed for these blocks separately. In other words, each block within a MB has a separate MV, and hence, a MB can have up to 16 MVs, depending upon the MB mode.

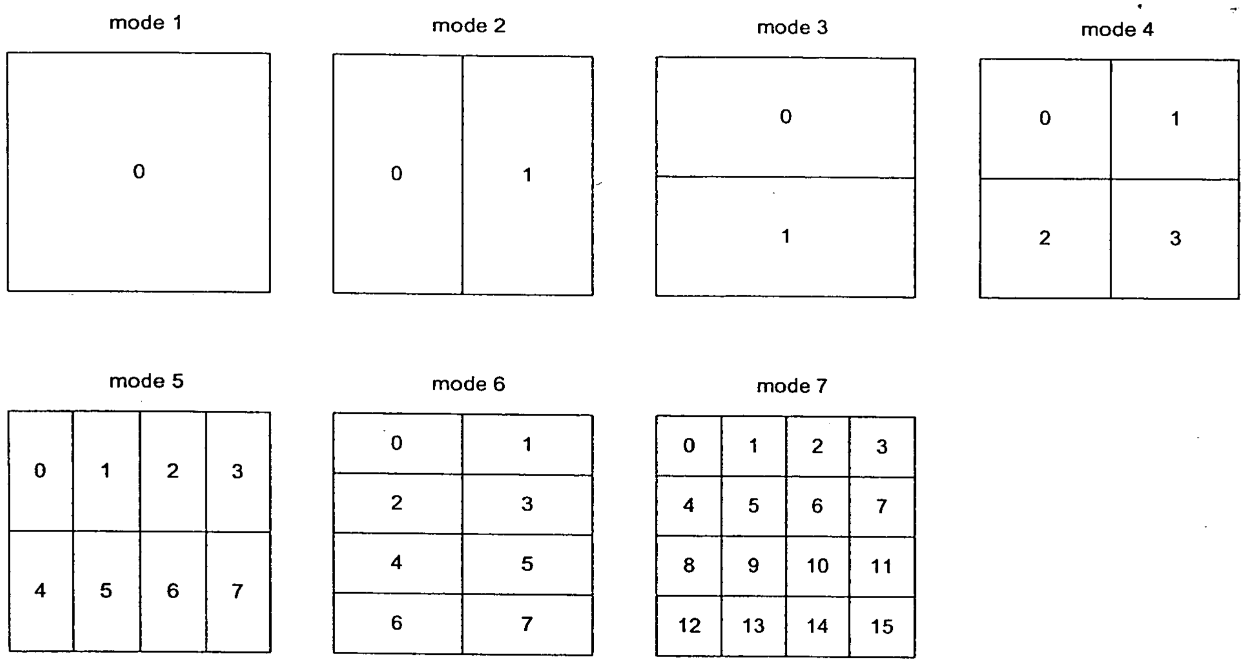

Fig. 1. Seven frame-based block patterns (modes) per MB.

The current TML 8.5 is designed for progressively scanned video materials. All the code modes in the current TML 8.5 are frame-based. Nature of interlace video materials and its possible impact on H.26L has been under discussion [2], [3], [4], [5], [6]. In particular, a frame of interlace video sequence consists of two fields, scanned at different time instants. Therefore, a video sequence consisting of interlaced video can be compressed in many different ways. They can be grouped in two main categories:

(a) Fixed Frame/Field and
(b) Adaptive Frame/Field

In Fixed Field/Frame method, an interlaced video sequence is always coded with either Field or Frame structure irrespective of the content. In Adaptive Field/Frame methods, depending upon the content, one adaptively selects whether to use the field or frame structure. Adaptation can be done at either Picture level or Macroblock level. This document provides results obtained by different coding tools in these categories. It is shown that if either fixed field or fixed frame structure is picked to compress interlaced video, it will lead to coding inefficiency as for some video sequences field structure is better and for others frame structure is better. Content adaptive Field/Frame based tools are introduced where a decision regarding field or frame structure is made based on the content type. It is shown that they perform better than fixed frame or field method by adaptively picking the optimal structure for a given content.

2. Interlace Coding

Inputs to the H.26L video code can be considered as a series of units at three different levels: sequence, picture, and MB. Frame or field coding can also be made adaptive at these three levels, that is, adaptive frame/field coding at sequence level, adaptive frame/field coding at picture level, and adaptive frame/field coding at MB level.

2.1 Fixed Frame/Field – Sequence Level

In this approach, all the frames of a sequence are encoded in either frame or field mode. The code mode of frame or field will not be changed during entire sequence. Additional changes to TML 8.5 are therefore minor, that is,

1. The sequence header needs to indicate whether frame or field mode is used.
2. For field mode, the code numbers assigned for field-based references stored in reference frame buffer are slightly different from for frame-based references. Specifically, the code numbers of 0, 1, 2, 3, …, are grouped into pairs of (0,1), (2,3), (4,5), …. These code number pairs are assigned to the reference frames of two fields according to their distances to the current frame. For each reference frame, the field of the same parity as the current field is given the smaller number of the code number pair assigned for the frame.
3. For field coding, the skipped MB will be reconstructed by copying the co-located MB in the most recently coded (past) I or P frame in the same field parity.

2.2 Adaptive Frame/Field Coding – Picture Level

For the picture level adaptation of frame/field, an input frame of a sequence can be encoded as one frame or two fields. Specifically,

1. An I frame can be coded as one I frame, or two I fields, or one I field and one P field,
2. A P frame can be coded as one P frame, or two P fields, or one P field and one B field, and
3. A B frame can be coded as one B frame, or two B fields.

In addition, for field-based B,

1. There can be multiple references for forward prediction, and the references can be any coded I or P field.
2. But, there is only one reference for backward prediction, and the reference is always the most recently coded future I or P of the same field parity.
3. For direct mode MB, the two MVs are calculated from the co-located MB in the future reference of the same field parity.

The changes to TML 8.5 therefore include

1. Picture header needs to indicate whether the frame is coded as one frame or two fields.
2. Two fields are coded sequentially. Hence, there will be only one picture header per frame of two fields.
3. Similar to at sequence level, for field mode, the code numbers for the field-based references in the reference frame buffer, 0, 1, 2, 3, …, are grouped into pairs of (0,1), (2,3), (4,5), …. The code number pairs are assigned to the reference frames of two fields according to their distances to the current frame. For each reference frame, the field of the same parity as the current field is given the smaller number of the code number pair assigned for the frame.
4. If field-based coding, the skipped MB will be reconstructed by copying the co-located MB in the most recently coded (past) I or P frame in the same field parity.

2.3 Adaptive Frame/Field Coding - MB Level

For MB level adaptation of frame/field coding, a MB can be coded in frame- or field-based. A frame/field flag may be required at MB level to indicate if the MB is coded in frame- or field-based, as shown in Fig. 2. "0" indicates frame-based coding and "1" field-based coding. Frame-based coding of a MB follows TML 8.5. Field-based coding will be described as follows.

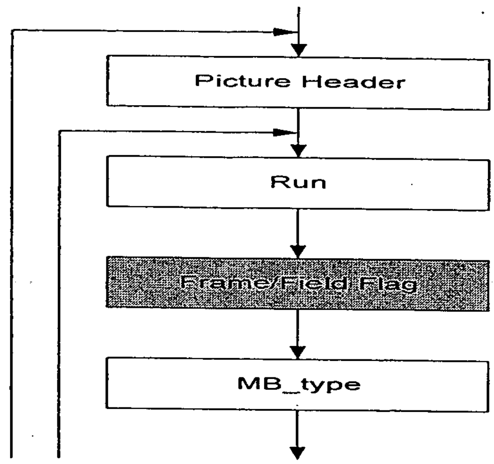

Fig. 2. A frame/field flag of one bit may be required to indicate if the following MB is coded in frame- or field-based.

Five additional block patterns are introduced for field-based coding, as shown in Fig. 3. For field-based coding, a MB is first split into two fields of top and bottom. The split MB is then further divided into one of five possible block patterns. Similar to in frame mode, a MB in field mode can be coded in intra or inter mode.

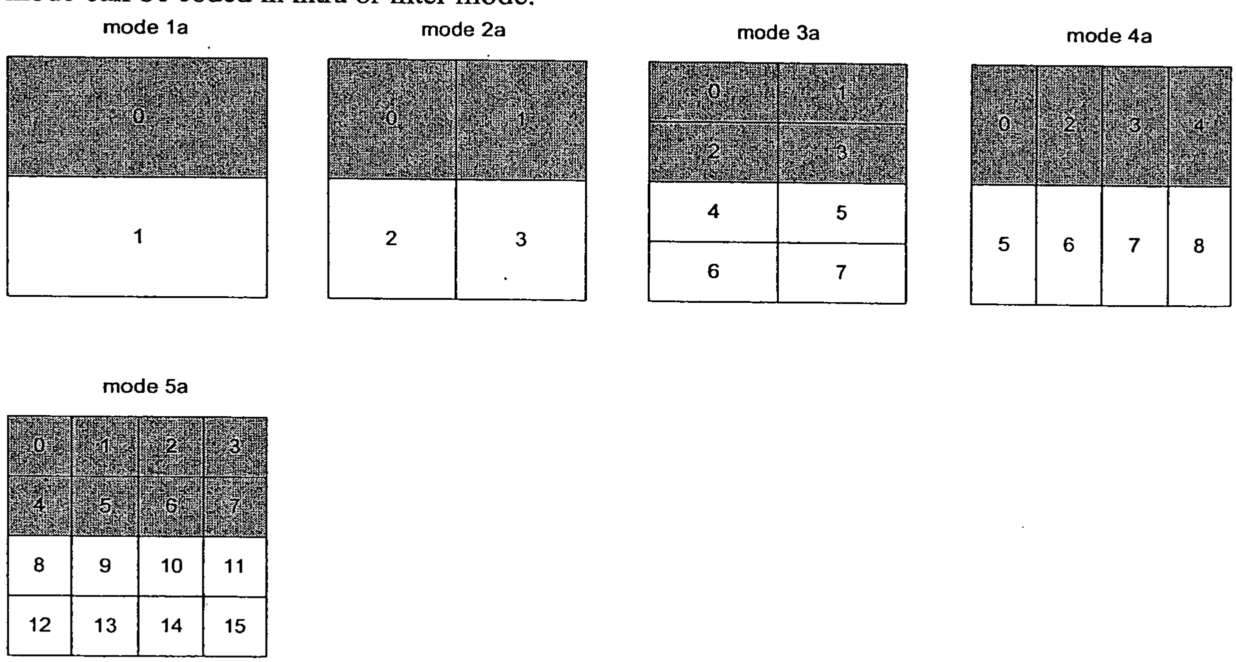

Fig.3. Five additional field-based block patterns (modes) per MB. The shadow (blue) areas are for top field and blank for bottom field.

Intra Prediction

For a MB in intra code mode, only mode 1, mode 7 and mode 5a are allowed, (see Fig. 1 and 3). Modes 1 and 7 are frame-based, and mode 5a is the field-based mode. Mode 5a is assigned the code number of 0.

Note that both mode 1 and mode 5a are coded with code number 0. But, there will be no confusion because of the frame/field flag at MB level (see Fig. 2).

For mode 5a, the neighboring pixels (A,B,C,D,E,F,G,H and I) of a block of 4x4 in calculating the prediction for the block (see Fig. 4) are the neighboring pixels in the same field parity. As example, Fig. 5 shows the neighboring pixels for the top-field blocks 0, 1, 2, ..., 7. Clearly, the neighboring pixels for blocks 5, 6 and 7 are in blocks 4, 0, 1, 2 and 3 in the same MB. For example, the neighboring pixels for block 7 are in block 2, 3, and 6. Blocks 2, 3, and 6 are in the same field parity as block 7. The left neighboring pixels for blocks 0 and 4 are in blocks 3 and 7 of the left MB in the same field parity. The above neighboring pixels for blocks 0, 1, 2 and 3 are in blocks 4, 5, 6 and 7 of the above MB in the same parity. Note that the left and above MB can be coded in any block patterns in either frame- or field-based.

I	A	B	C	D
E	a	b	c	d
F	e	f	g	h
G	i	j	k	l
H	m	n	o	p

Fig. 4. Neighboring pixels of a 4x4 block in mode 5a.

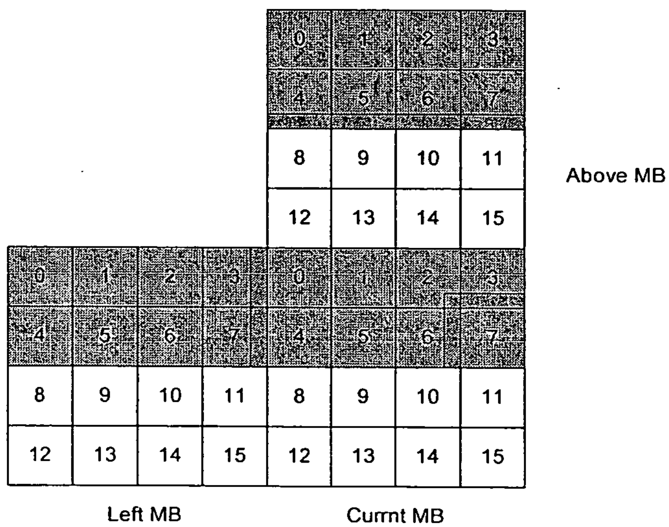

Fig. 5. Neighboring pixels of a block have to be in the same field parity. The light shadow (blue) areas are for top field and blank for bottom field. The dark shadow (green) areas are the neighboring pixels of a block on the right or blow.

Coding of Intra Prediction Modes

As in frame-based coding, the prediction mode of a 4x4 field-based block is coded based upon the prediction modes of the *neighboring* blocks. For interior block 5, 6, 7, 13, 14 or 15 of a field-based MB (Fig. 5), the neighboring blocks used in coding of intra prediction mode are simply the above and left blocks. For example, the neighboring blocks of block C in Fig. 6 are blocks A and B. For other boundary blocks (0, 1, 2, 3, 4, 8, 9, 10, 11, and 12), the neighboring blocks are defined as follows:

1. If the above or the left MB (Fig. 6) is also coded in field-based, as shown in Fig. 6a, the neighboring blocks of the boundary blocks in current MB are in the same field of the above or the

left MB. For example, the neighboring blocks of current blocks C' and C'' are A' and B', and A'' and B'', respectively.

2. If the above or the left MB is coded in frame-based, as shown in Fig. 6b, the neighboring blocks of the boundary blocks in the current MB are on the bottom or the left of the above or the left MB. For example, the neighboring blocks of current blocks C' and C'' are A' and B', and A'' and B', respectively.

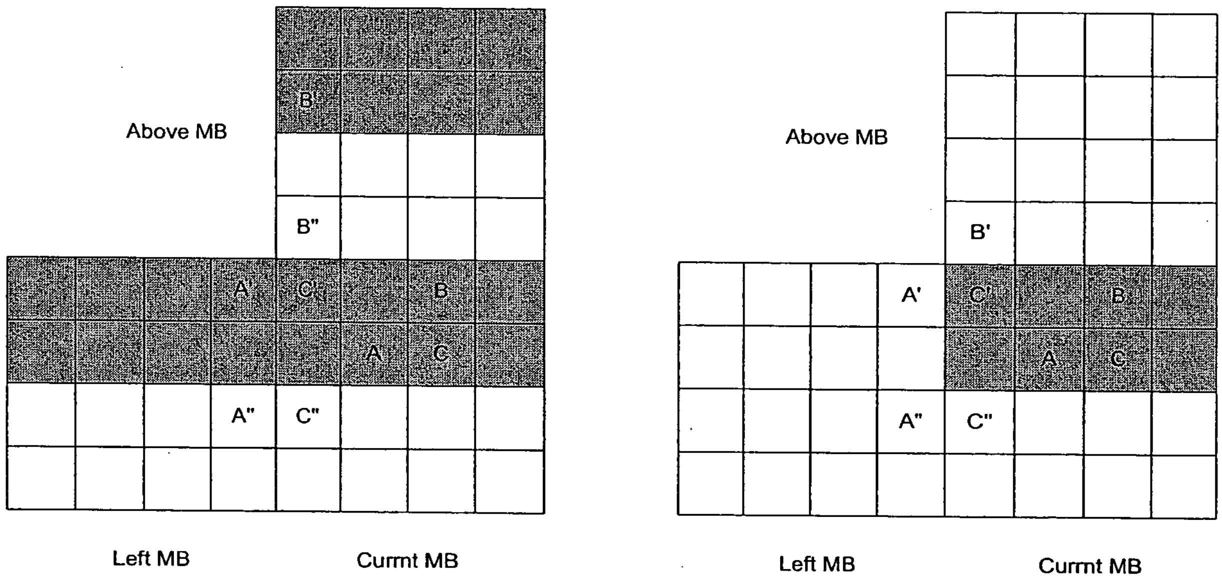

Fig. 6. Coding of intra prediction mode. 6a) the shadow areas are for top field and blank for bottom field. 6b) The shadow areas (blue) in current MB are for top field and blank for bottom field. The above and left MBs are frame-coded.

Prediction of Chroma

There is only one prediction mode for interlace chroma blocks, as shown in Fig. 7. A, B, C and D are four 8x8 chroma blocks of a MB. A, B are in the top field and C, D in the bottom field. S0, S1, S2, S3, S4, and S5 are the sums of the neighboring pixels of A, B, C and D in the same field parity. S0, S1, S2 are calculated from the top field and S3, S4, S5 from the bottom field.

If S0, S1, S2, S3, S4 and S5 are all inside the frame, the predictions for A, B, C and D are as follows
A=(S0+S2+4)/8
B=(S1+2)/4
C=(S3+S5+4)/8
D=(S4+2)/4

If only S0, S1, S3, and S4 are inside the frame, the predictions for A, B, C and D are follows
A=(S0+2)/4
B=(S1+2)/4
C=(S3+2)/4
D=(S4+2)/4

If only S2 and S5 are inside the frame, the predictions for A, B, C and D are as follows
A=(S2+2)/4
B=(S2+2)/4
C=(S5+2)/4
D=(S5+2)/4

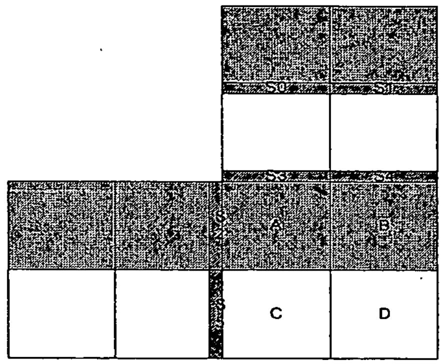

Fig. 7. The prediction of interlace chroma blocks. The shadow areas (blue) are for top field and blank for bottom field.

Chroma Residual Coding

If the luminance component is coded in field-based, the chroma component will follow. The chroma will be split into the top and the bottom fields, as shown in Fig. 8. 4x4 transform is performed on each block of 4x4. The DC coefficients are then grouped into two 2x2 blocks. 2x2 transform is performed on the 2 2x2 DC blocks.

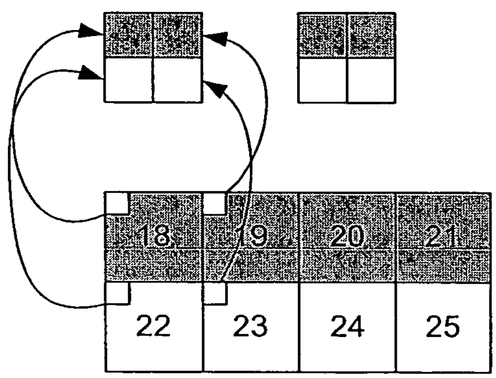

Fig. 8. Field-based chroma residual coding. The shadow areas (blue) are for top field and blank for bottom field.

Inter Code Mode

For inter code modes, ME/MC is preformed on each frame/field block. As in frame mode, there are also up to 16 blocks per MB in field mode (see Fig. 3). The smallest block is 4x4. Hence, there are up to 16 MVs per MB in field mode. The reference fields for a field-based block can be either the top or the bottom fields of any previously coded frames. Six field-based ME/MC code modes can be the possible code modes in encoding a MB. They are field inter 16x16, inter 8x16, inter 8x8, inter 4x8, inter 8x4, and inter 4x4. The field inter 16x16 is similar to the inter 8x16 mode except the motion vectors for the top field and the bottom field are forced to be the same, and the reference field for the bottom field is the field immediately following the reference field for the top field. The advantage of the field 16x16 mode lies in the fact that it reduces the number of bits required to send the motion information and the extra reference field information. These code modes are assigned the code numbers of 0, 1, 2, 3, 4 and 5 accordingly. There will be no

confusion with frame-based inter code modes because of the frame/field flag at MB level (see Fig. 2).

Reference Frames/Fields
Multiple reference frames are allowed for temporal prediction. The reference frame for a MB has to be indicated in the syntax at the MB level. TML8 (frame-based) assigns code number 0 to the most recently coded frame, 1 to the second most recently coded frame, and so on. This implies the closer the reference frame to the current frame, the more likely the reference frame will be selected, and hence the smaller the code number is assigned. For field-based coding, the reference can be either the top or the bottom field. The reference field will also be indicated in the syntax at the MB level. Similar to at both sequence and picture levels, the code number for reference fields are arranged as follows:

1. The code numbers for the field-based references stored in the reference buffer, 0, 1, 2, 3, …, are grouped into pairs of (0,1), (2,3), (4,5), …
2. The code number pairs are assigned to the reference frames of two fields according to their distances to the current frame.
3. For each reference frame, the field of the same parity as the current field is given the smaller number of the code number pair assigned for the frame.

Both the encoder and the decoder follow this rule.

In addition, each field MB is treated as an independent MB in the sense that it can have its own reference field for all blocks in that field MB (these two reference fields are not necessarily come from the same frame). Hence, two codes are put in the syntax at the MB level to indicate reference field of its top and bottom field when field MC is employed.

PMV for Inter Blocks
The MVs are differentially coded. Assume block E, as shown in Fig. 9, is inter coded. The prediction for MV of block E is the median of the MVs of the neighboring blocks A, B, C and D, as shown in Fig. 9. For definition of positions of A, B, C, and D, refer to TML8 [1]. Block E can now be either frame- or field-based block. In other words, all the frame- and field-based modes are allowed.

1. If E is in frame-based, the MVs of A, B, C and D used in calculating PMV are also frame-based. If block A, B, C, or D is coded in field-based, the two field-based MVs are averaged and the vertical component of this averaged MV is multiplied by 2.
2. If E is in field-based, the MVs of A, B, C and D used in calculating PMV are also field-based in the same field parity. If A, B, C, or D is frame coded, the vertical component of MV is divided by 2.

For directional segmentation, follow the same conventions as in TML8 [1], but the neighboring blocks are in the same field parity.

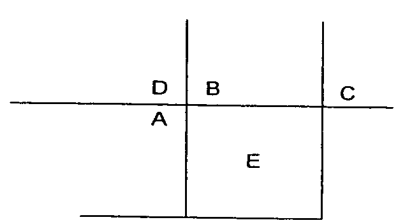

Fig. 9. Calculation of PMV.

3. Computer Simulations

To evaluate the additional interlace tools for H.26L video coding, computer simulations were carried out for five test video sequences: *Bus*, *CheerLeader*, *Coastguard*, *Mobile* and *Stefan*. All the test sequences have a spatial resolution of 352x480 (HHR) with a color-sampling ratio of 4:2:0, interlaced at a frame rate of 30 frames/sec. The following picture structures are tested:

1. I picture only: I I I I I ,...
   a. Fixed frame coding.
   b. Fixed field coding.
   c. Adaptive frame/field coding at picture level
   d. Adaptive frame/field coding at MB level
2. I and P pictures only: I P P P P, ...
   a. Fixed frame coding.
   b. Fixed field coding. The first I frame is coded as one I field and one P field.
   c. Adaptive frame/field coding at picture level. If the first I frame is coded in field-based, the first field of this I frame is coded as I and the second as P, that is, IP PP PP PP, ...
   d. Adaptive frame/field coding at MB level.
3. I, P and B pictures: I B B P B B P, ...
   a. Fixed frame coding.
   b. Fixed field coding. The first I frame is coded as one I field and one P field
   c. Adaptive frame/field coding at picture level. If the first I frame is coded in field-based, the first field of this I frame is coded as I and the second field as P, that is, IP BB BB PP BB BB PP,....
   d. Adaptive frame/field coding at MB level

Note that the frame-based coding is the same as TML 8.5. For the picture level adaptation, the criterion used for selecting either frame or field coding is RD-based. The cost function is defined as

$$\cos t = Distortion + \lambda BitRate\,.$$

Table 1 shows the test conditions.

Table 1. Test conditions

Entropy C.	MC	Hadamard	Ref. Frames	Search Res.	RD Opt.	Search R.
UVLC	1/4 pel	Yes	3	2	yes	16

3.1 Results for I Picture only

For I picture only, four fixed QPs are used for the entire sequence. They are 19, 22, 25, and 28. Fig. 10-14 show the PSNR with respect to bit rate for the five test sequences with adaptation of frame/field coding at sequence, picture and MB levels. All the frames are coded as I picture. The curves with square marks are for fixed frame coding for the entire sequence, the curves with diamond marks for fixed field coding for the entire sequence, the curves with triangle marks for the picture level adaptation, and the curves with "x" marks for the MB level adaptation.

Four of five test sequences (*CheerLeader*, *Coastguard*, *Mobile* and *Stefan*) demonstrate better results with fixed frame coding and one (*Bus*) better with fixed field coding. The difference in PSNR can be up to more than 2 dB (*Mobile*). This implies that the fixed decision of frame or field coding is not be a good solution because frame-based coding may be good for some sequences, but inefficient for others. Similarly, field-based decision is good for some sequences and inefficient for others. Also, for a sequence consisting of mixed frame types (e.g. some of Bus type and other of Mobile type) fixed field or frame decision will be good for some video frames of the sequence, but inefficient for other.

The picture level adaptation of frame/field coding (curve with triangle marks) is either equal to, or better than, the frame- and field-based coding. Fig. 15 shows the number of frames coded in field with the picture level adaptation of frame/field coding. As seen, *Bus* sequence is 100% coded in field while *Cheer* and *Mobile* 100% in frame. The picture level adaptation therefore matches the field-based coding for *Bus* and the frame-based coding for *Cheer* and *Mobile*. For *Coastguard*, 5~6 frames out of 150 frames are coded in field and for *Stefan*, 45~56 frames of 150 frames in field. The picture level adaptation is hence better than both frame- and field-based coding for both *Coastguard* and *Stefan*. The picture level adaptation of frame/field coding is a much better solution for interlace video materials than frame- and field-based coding for I pictures.

The MB level adaptation of frame/field shows further improvement over the picture level adaptation for three (*CheerLeader*, *Cheer* and *Stefan*) out of the five test sequences by up to 0.4 dB (*Cheer*). For other two sequences (*Coastguard* and *Mobile*), the MB level adaptation is equally good as the picture level adaptation of frame/field.

3.2 Results for I and P Pictures only

For I and P pictures only, four fixed QPs are used for the entire sequence. They are 19, 22, 25, and 28. Fig. 16-20 show the PSNR with respect to bit rate for the five test sequences with adaptation of frame/field coding at sequence, picture and MB levels. The first frame is coded as either one I frame for frame-based coding, or one I field and one P field for field-based coding. The rest frames are all coded as P. The curves with square marks are for fixed frame coding for the entire sequence, the curves with diamond marks for fixed field coding for the entire sequence, the curves with triangle marks for the picture level adaptation, and the curves with "x" marks for the MB level adaptation.

Two of five test sequences (*Coastguard* and *Mobile*) demonstrate better results with frame-based coding and other three (*Bus, CheerLeader* and *Stefan*) with field-based coding. The difference in PSNR can be up to more 3 dB (*Stefan*). This again implies that the fixed decision of frame or field is not be a good solution because frame-based coding performs better than field-based coding for some sequences, and worse for some other sequences.

The picture level adaptation of frame/field coding (curve with triangle marks) performs better than both the frame- and field-based coding. Fig. 21 shows the number of field-coded frames with the picture level adaptation of frame/field coding. As seen, the majority of frames for *Bus* CheerLeader, and Stefan are field-coded. Hence, the picture level adaptation for theses three sequences is close to, but slightly better than, the field-based coding. On the other hand, more frames of *Coastguard* and *Mobile* are frame-coded. Hence, the picture level adaptation is close to, but better than, the frame-based coding for *Coastguard* and *Mobile*. It is again shown that the picture level adaptation of frame/field coding is a much better solution for interlace video materials than fixed frame- and field-based coding for I and P pictures.

The MB level adaptation of frame/field shows further improvement over the picture level adaptation for *Mobile*. For some other sequences such as *Bus* and *Coastguard* the performance of adaptive MB level frame/field algorithm is quite similar to that of the adaptive picture level algorithm. Further work is in progress to improve the performance of MB level frame/field adaptation algorithm and to generate more results.

3.3 Results for I, P and B

For I, P and B pictures, four fixed QPs of 19, 22, 25 and 28 are used for the entire sequence. Fig. 22-26 show the PSNR with respect to bit rate for the five test sequences with adaptation of frame/field coding at sequence level. The first frame is coded as either one I frame for frame-based coding, or one I field and one P field for field-based coding. The rest frames are coded as either P or B. There are two B frames between P frames. In field mode, a P frame is coded as two P fields, and a B frame is coded as two B

fields. The curves with square marks are for fixed frame coding for the entire sequence, and the curves with diamond marks for fixed field coding for the entire sequence.

Frame-based coding is seen to give better results than field-based coding for *Mobile* by about 1 dB. Field-based coding demonstrates better results for *Bus* and *Stefan* by up to 2 dB (*Stefan*). It is again approved that it is necessary to adapt frame and field coding based upon the content.

Simulations on MB and picture levels are still in progress. The results may be presented at the meeting.

4. Conclusions

The current TML 8.5 is designed for progressively scanned video materials only. Field-based coding can improve the coding performance for interlace video materials. Frame/field coding can be done in two ways: fixed frame or field and adaptive frame/field. Adaptation can be done at two levels: picture level and MB level. The changes to TML 8.5 on frame and field coding for interlace video are described.

The simulation results demonstrated that

1. Fixed frame or fixed field may be good for some sequences, but not for others. Hence, fixed frame/field coding may not be the solution for interlace video materials.
2. Performance of picture level adaptation is always better than, or equal to, both fixed frame and fixed field coding.
3. Adaptation of frame/field coding at MB level can further improve the performance for some sequences.

References:

1. ITU-T, Video Coding Expert Group (VCEG), "H.26L test model long term number 8 (TML-8) draft", 6/28/2001.
2. A. Luthra, R. Gandhi, K. McKoen, Y. Yu, K. Panusopone, D. Baylon, L. Wang, "Performance of MPEG-4 Profiles Used for Streaming Video and Comparison with H.26L," ISO/IEC JTC1/SC29/WG11 M7227, Sydney, July 2001.
3. P. Borgwardt, "Handling Interlaced Video in H.26L ," ITU-T VCEG-N57, Santa Barbara, Sept 2001.
4. C. Fogg, "Adaptive Field/Frame Block Coding Experiment Proposal," ITU-T VCEG-N76
5. M. Gallant, G. Cote, "High Rate, High Resolution Video using H.26L," ITU-T VCEG-N84, Santa Barbara, Sept 2001.
6. P. Borgwardt (Activity Coordinator), "Core Experiments on Interlaced Video Coding," ITU-T VCEG-N85, Santa Barabara, Sept 2001.

Patent Statement

Motorola Broadband Communications Sector may have intellectual property rights related to this contribution to ITU-T Q.6/SG16 Video Coding Experts Group, and, within the scope of a final published standard within this Group, would subscribe to sub clause 2.2 of the ITU-T patent policy under the condition of reciprocity from other contributors.

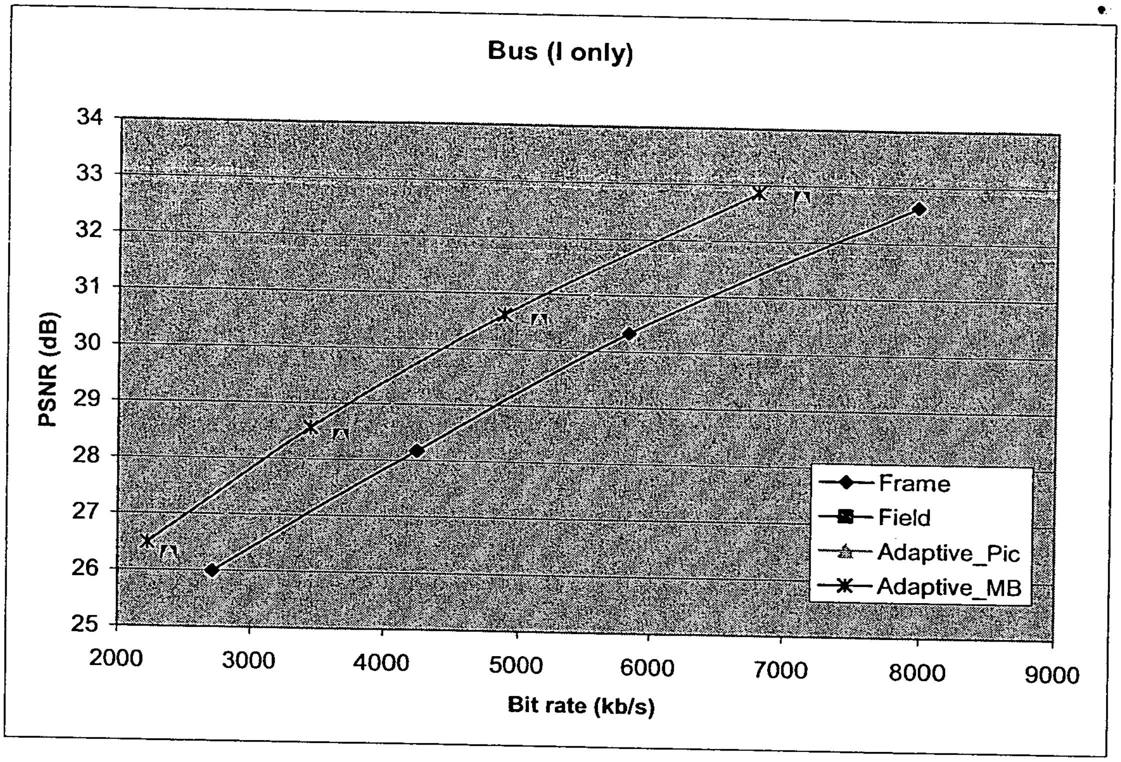

Fig. 10. PSNR vs bit rate curve for the *Bus* sequence.

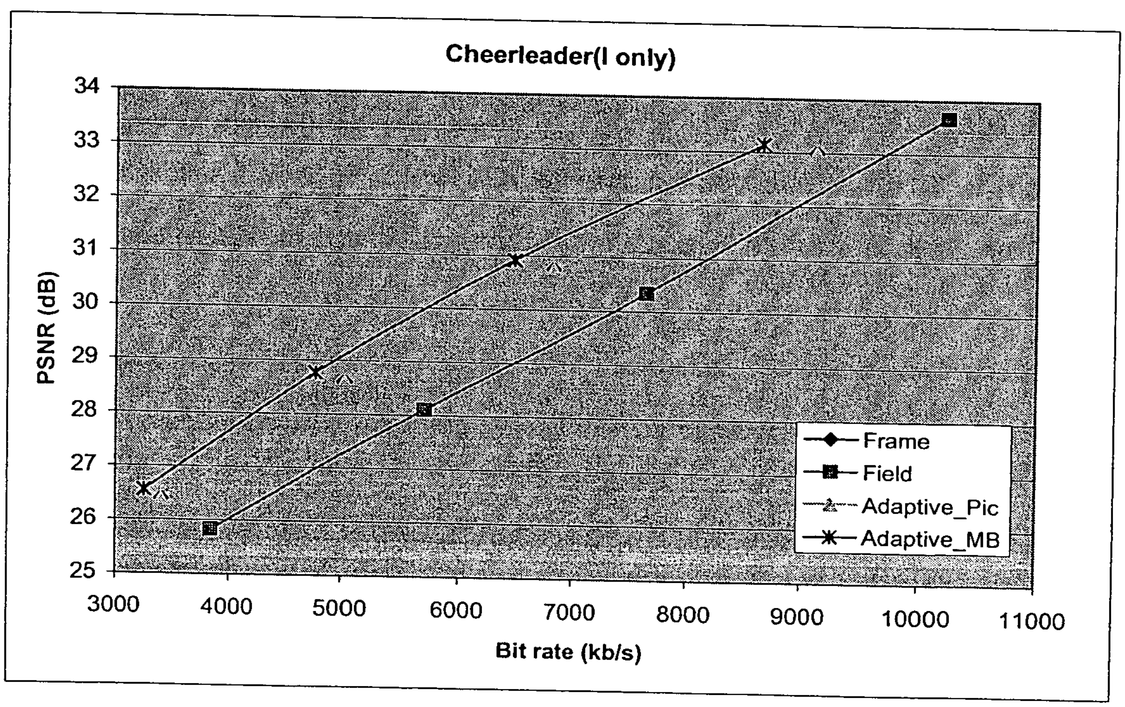

Fig. 11. PSNR vs bit rate curve for the *Cheerleader* sequence.

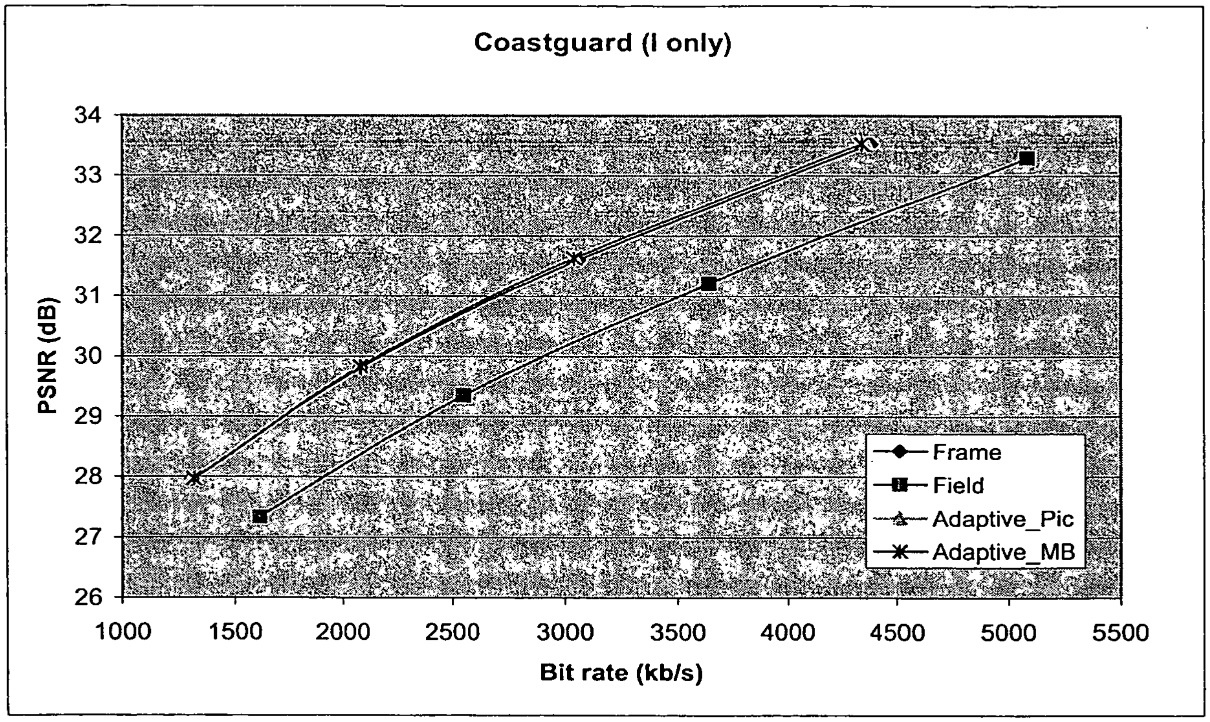

Fig. 12. PSNR vs bit rate curve for the *Coastguard* sequence.

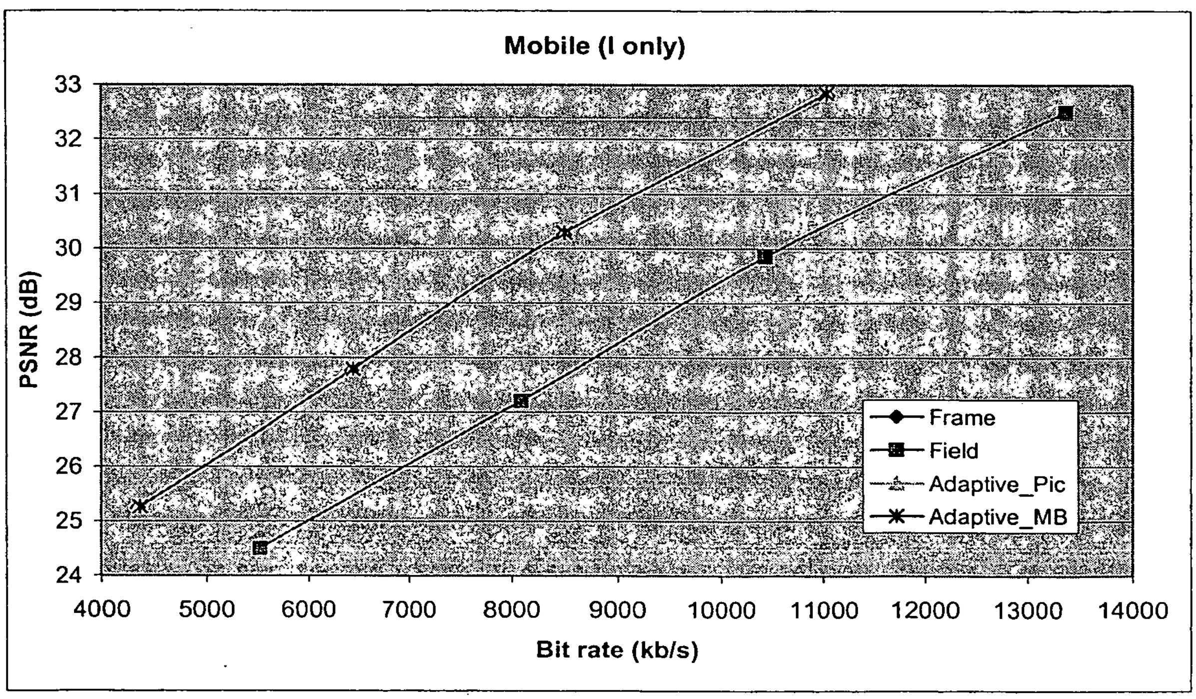

Fig. 13. PSNR vs bit rate curve for the *Mobile* sequence.

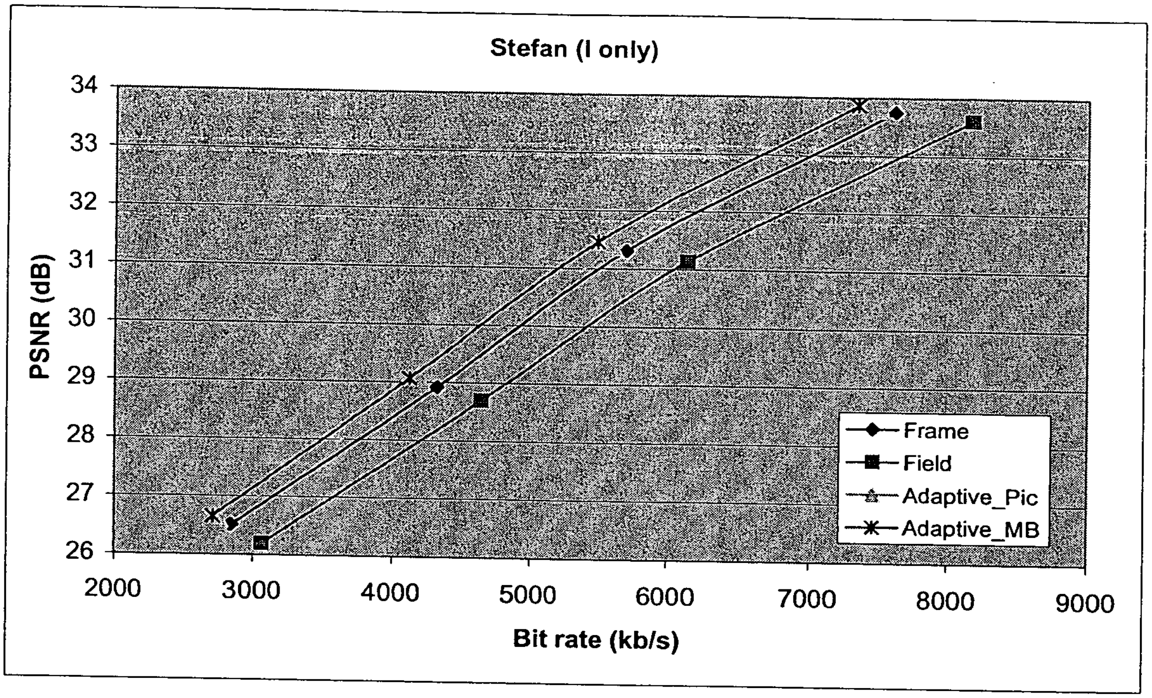

Fig. 14. PSNR vs bit rate curve for the *Stefan* sequence.

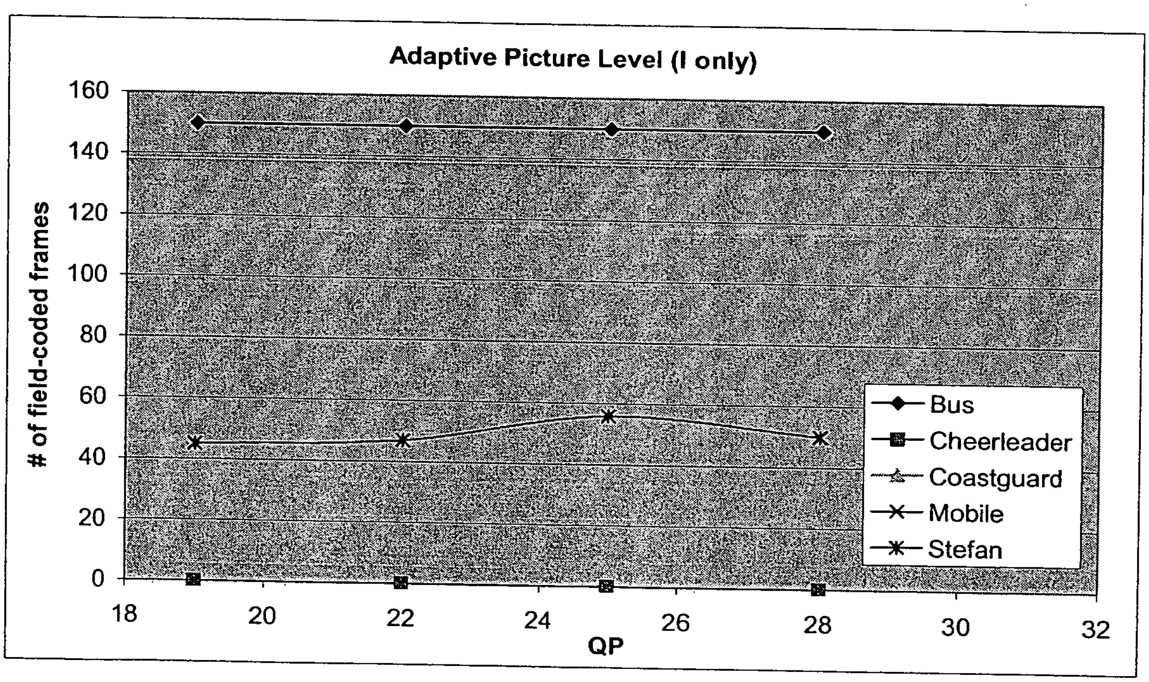

Fig. 15. Number of field coded pictures as a function of the QP value.

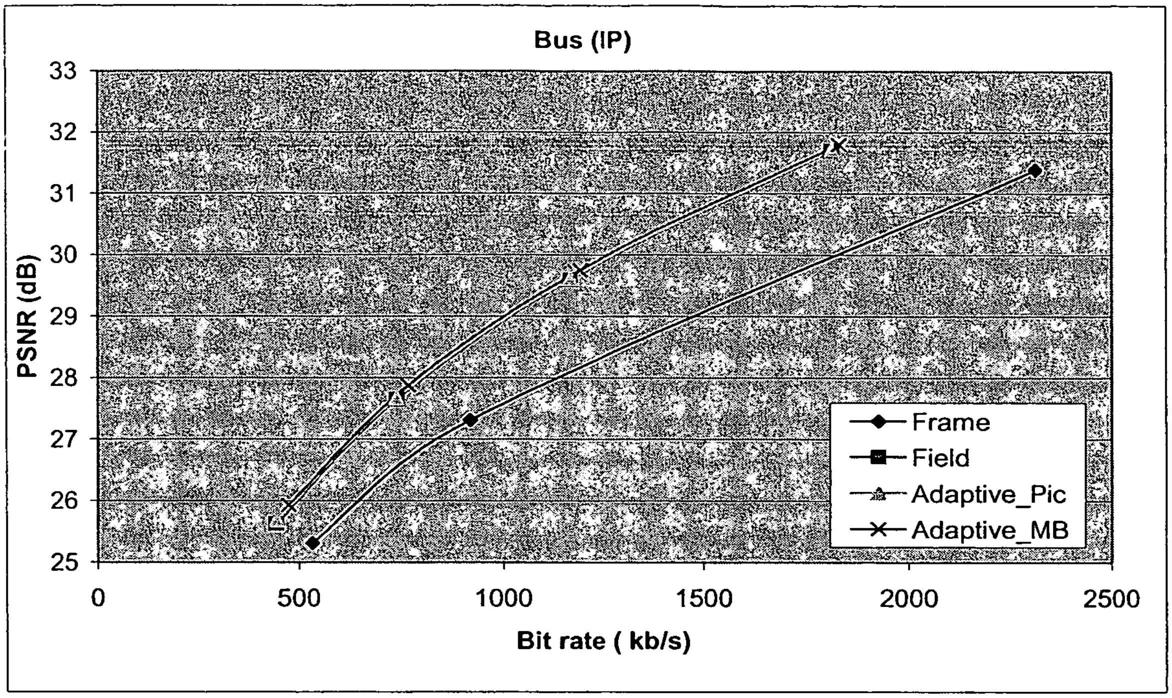

Fig. 16. PSNR vs bit rate curve for the *Bus* sequence.

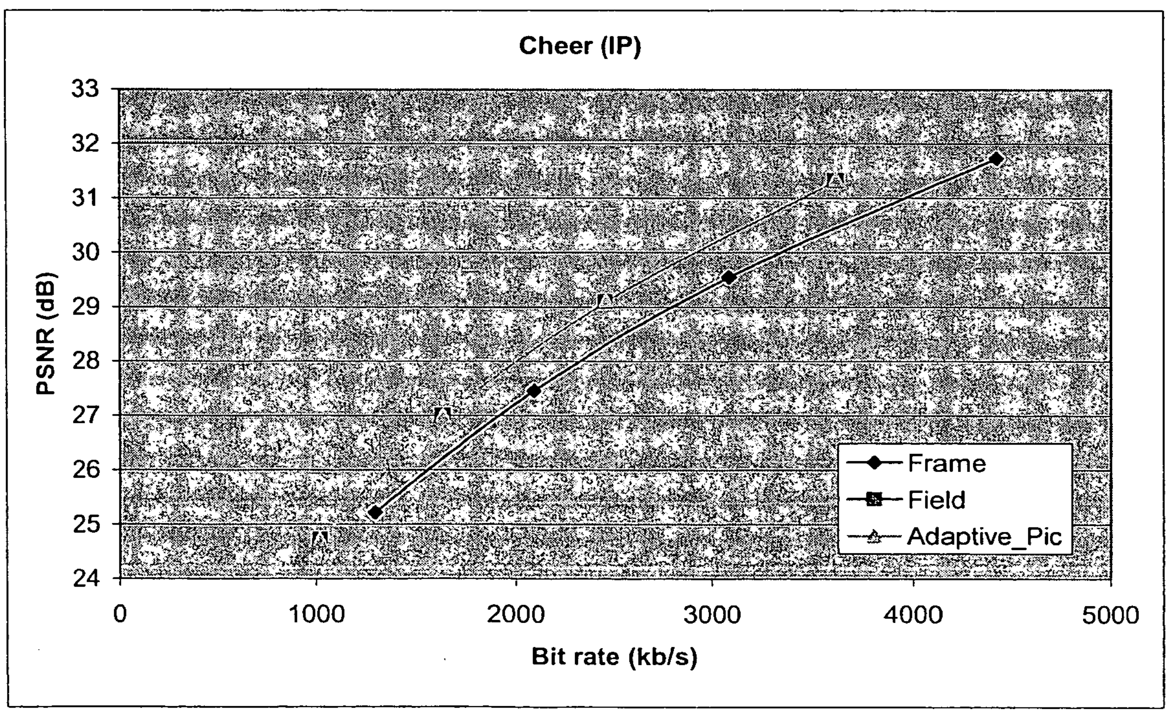

Fig. 17. PSNR vs bit rate curve for the *Cheerleader* sequence.

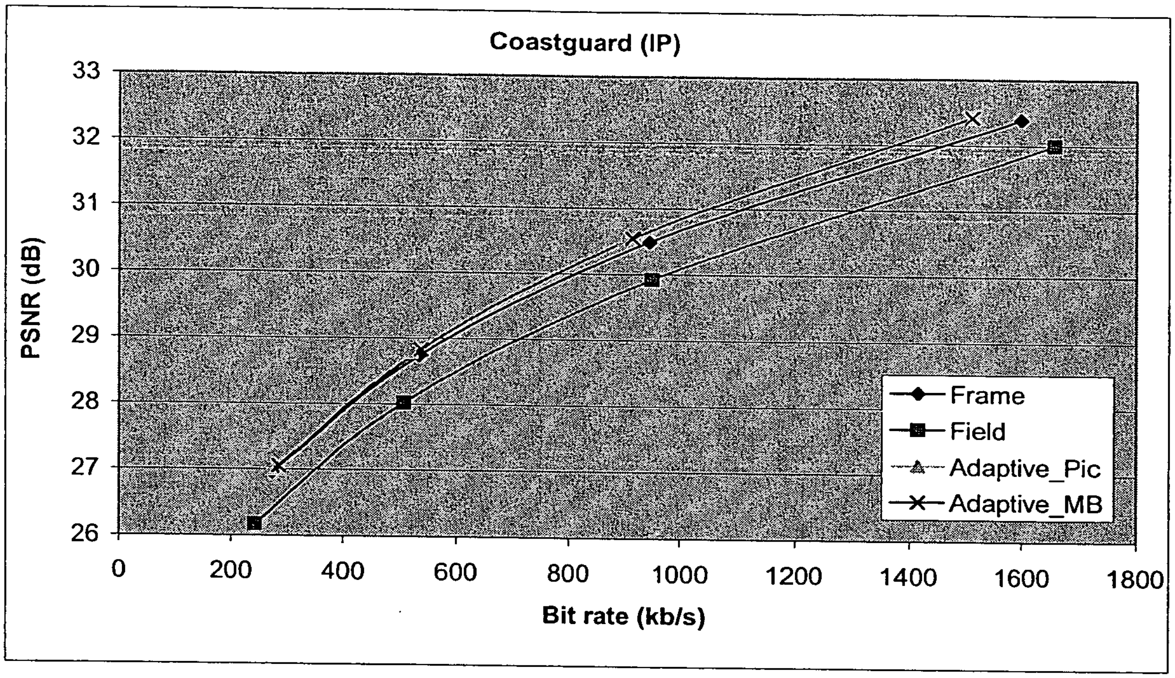

Fig. 18. PSNR vs bit rate curve for the *Coastguard* sequence.

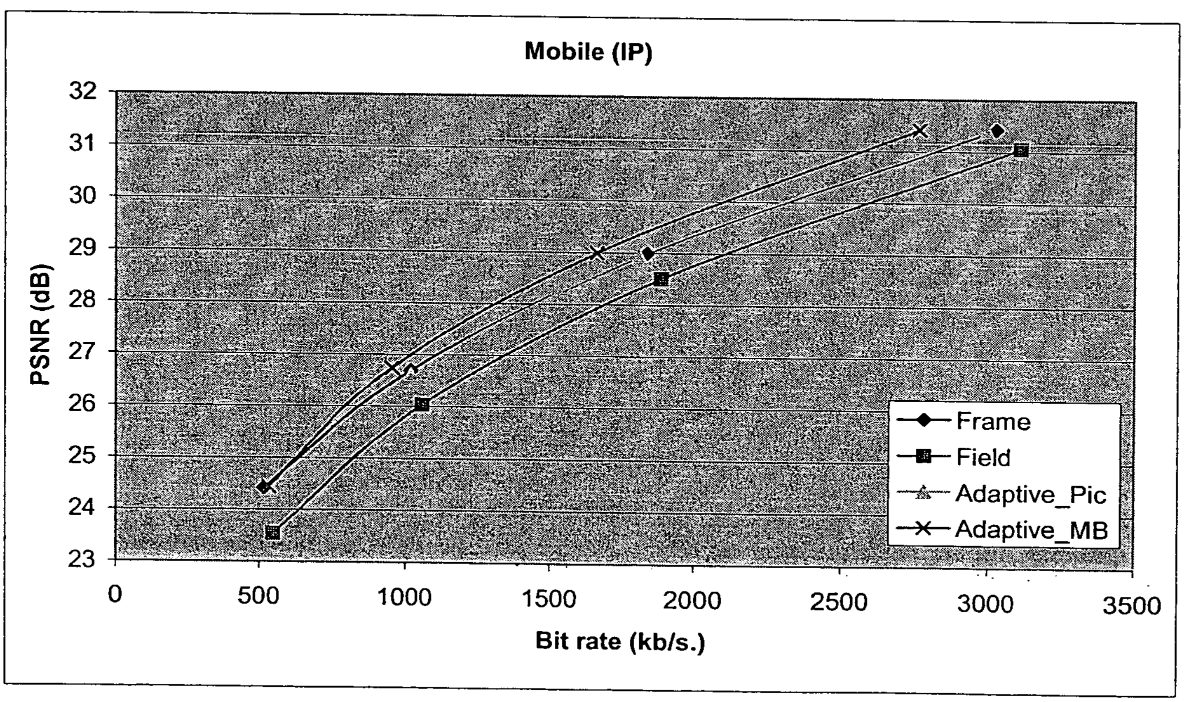

Fig. 19. PSNR vs bit rate curve for the *Mobile* sequence.

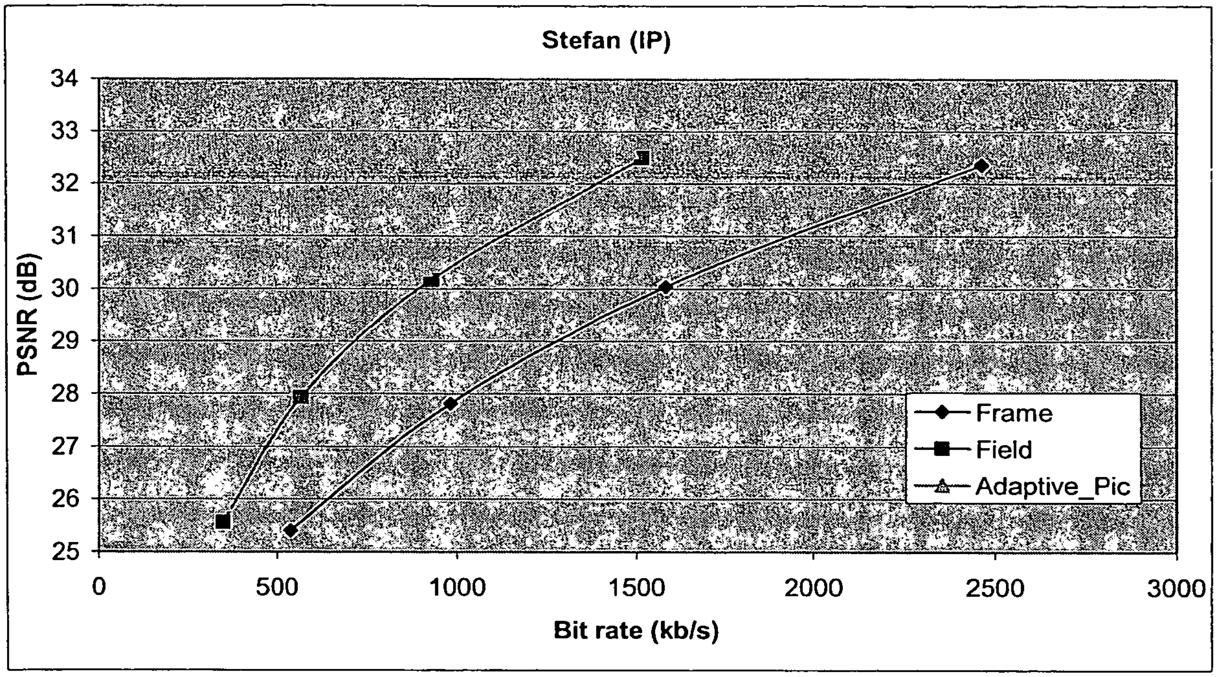

Fig. 20. PSNR vs bit rate curve for the *Stefan* sequence.

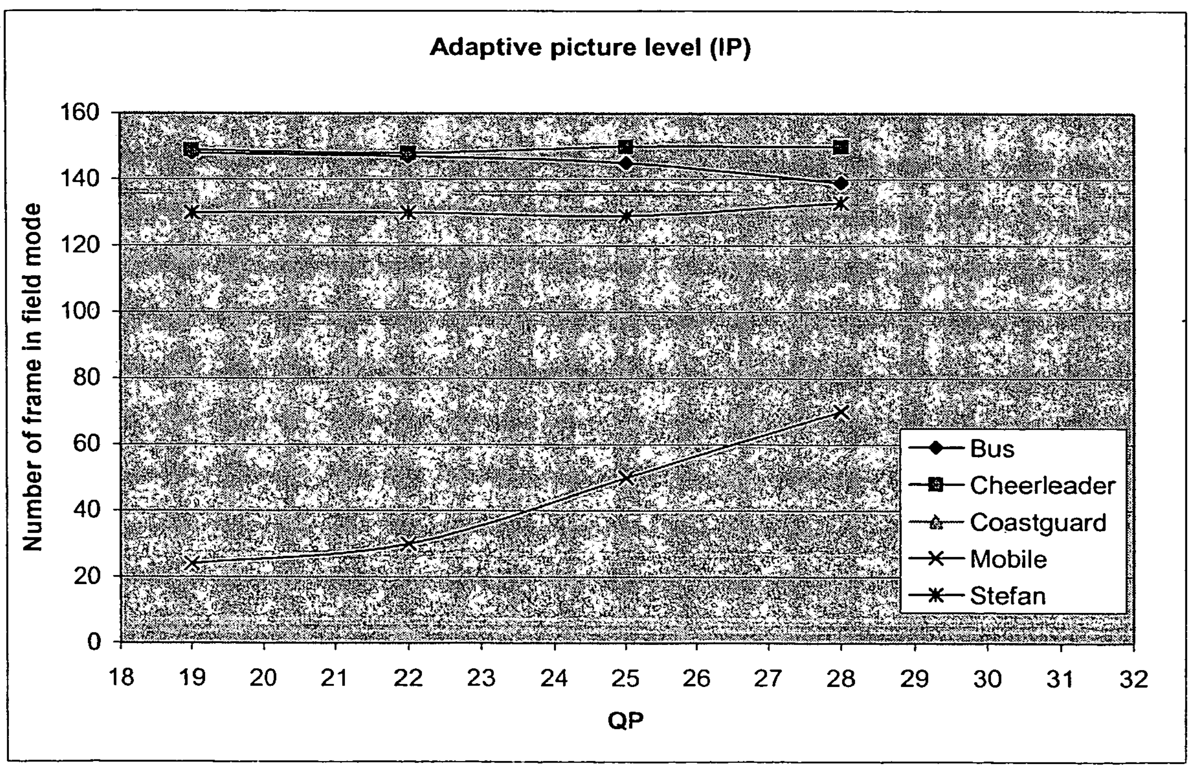

Fig. 21. Number of field coded pictures as a function of the QP value.

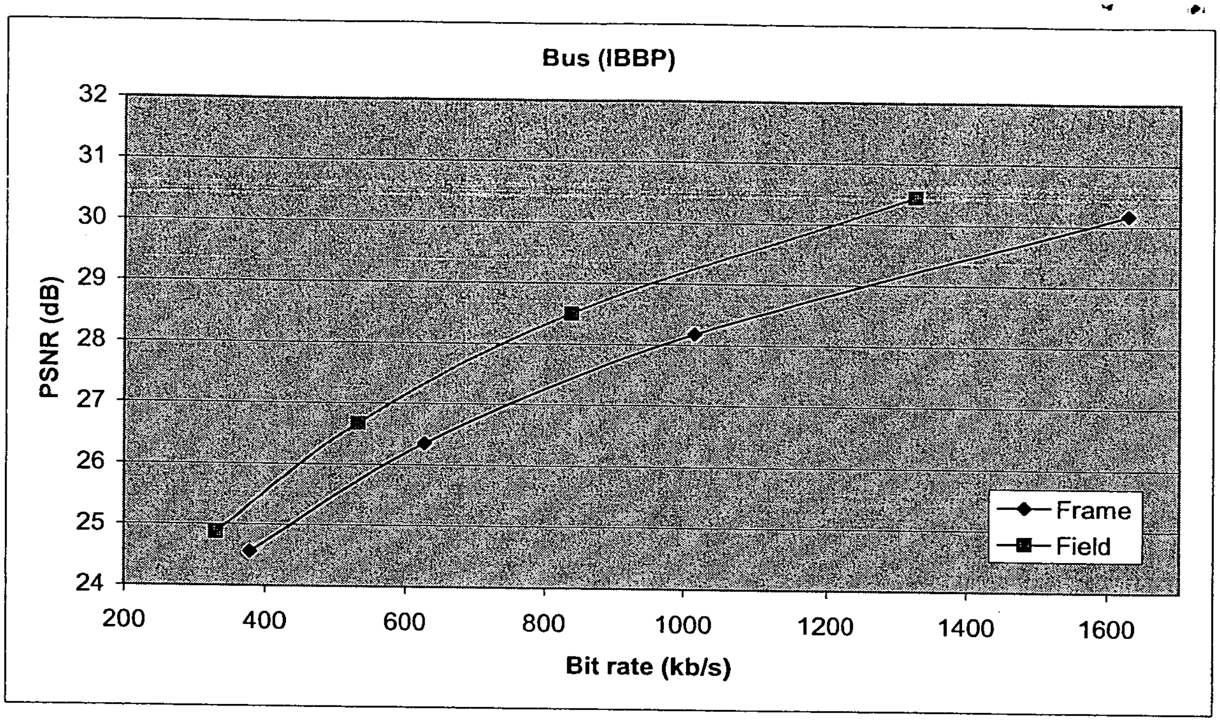

Fig. 22. PSNR vs bit rate curve for the *Bus* sequence.

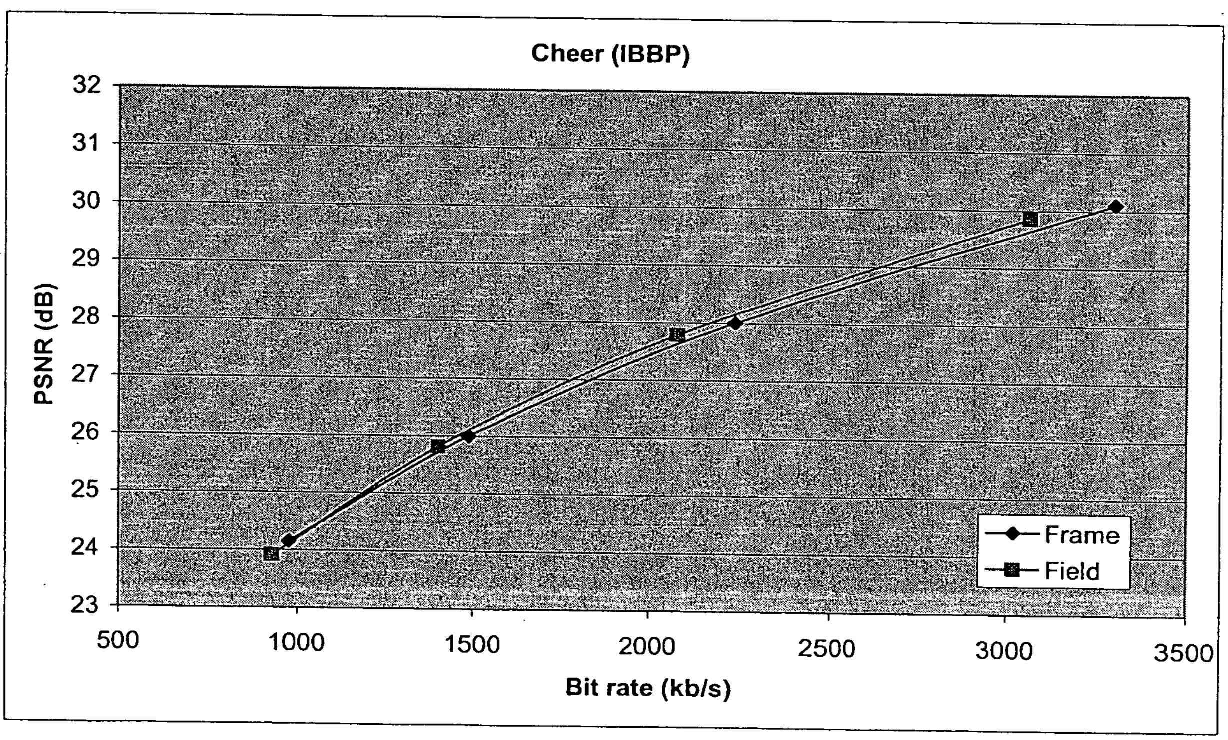

Fig. 23. PSNR vs bit rate curve for the *Cheerleader* sequence.

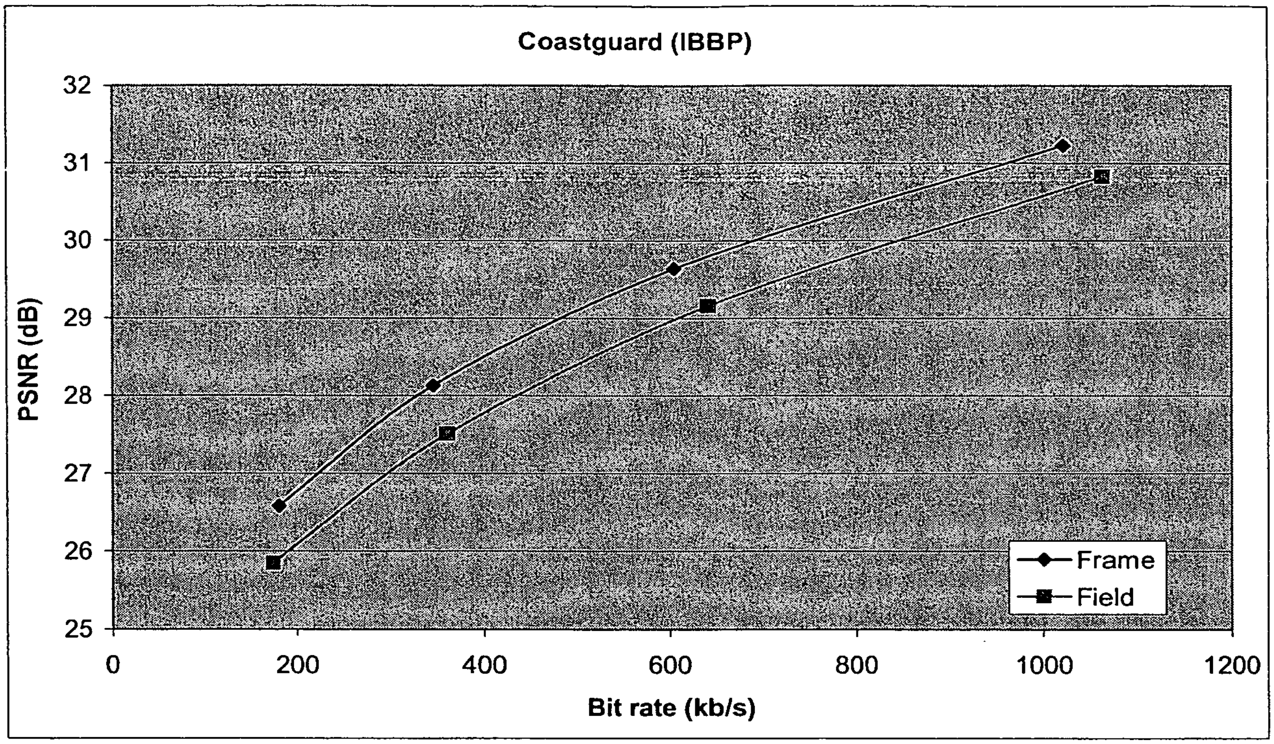

Fig. 24. PSNR vs bit rate curve for the *Coastguard* sequence.

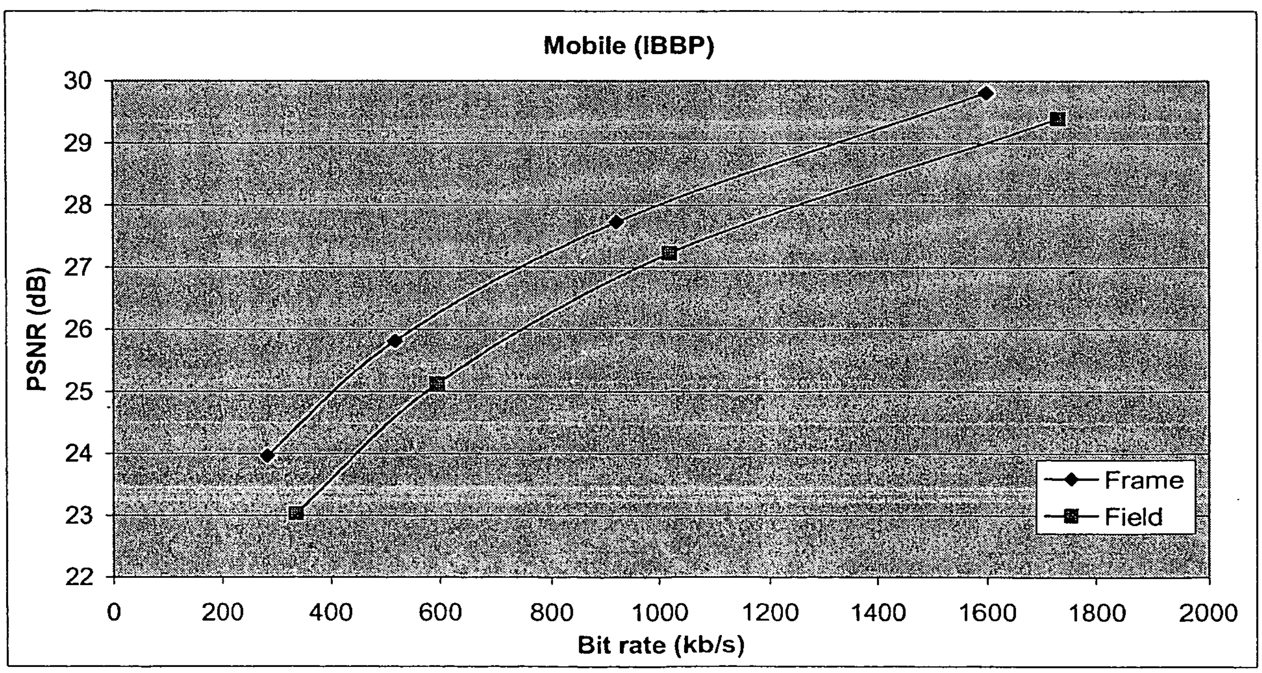

Fig. 25. PSNR vs bit rate curve for the *Mobile* sequence.

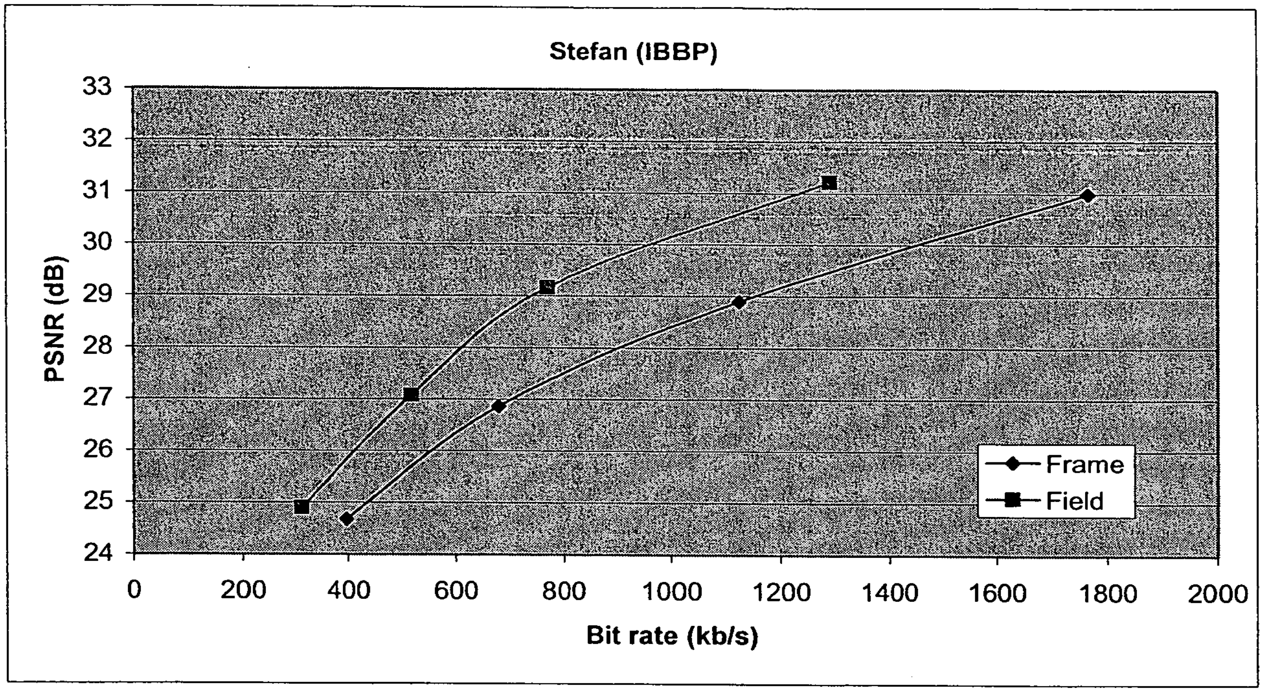

Fig. 26. PSNR vs bit rate curve for the *Stefan* sequence.